An Explicit Probabilistic Model For Genome Compartmentalization

Hagai Kariti, Tal Feld, Noam Kaplan

Department of Physiology, Biophysics & Systems Biology, Rappaport Faculty of Medicine, Technion – Israel Institute of Technology

Our DNA sequence stores important information. The physical spatial organization of DNA is associated with how this information is regulated. In the past decade, an experimental method called Hi-C has revealed several genomic structures. Hi-C produces a matrix of interaction frequencies between every two points in the genome that contains multiple interaction patterns. A major goal is understanding how these patterns emerge and their biological function. Currently, most analysis methods are based on data mining heuristics, which cannot make predictions, or on machine learning methods which are difficult to interpret. Here we present a new approach, using explicit probabilistic models based on hypotheses about the underlying mechanisms of the system. We focus on a pattern called genomic compartments. We show that our model outperforms current heuristics for extracting this signal from sparse or noisy inputs. We then use our modelling framework to gain biological insights not possible with current methods.